95-G-034.

Final Report:

Speech Analysis of ATCs

Philip Lieberman and Emily Pickett

Brown University

October 1, 1996

1. Introduction

The objective of this study is to identify acoustic properties of air traffic controller (ATC) communications associated with changes in workload. The investigation of acoustic correlates of perceived emotional stress and cognitive load is an active field of inquiry. While researchers have not yet identified consistently reliable quantifiable factors, several aspects of speech production have been shown to be related to physiological and task-induced stress (e.g., Lieberman et al, 1995; Absil et al, 1995; Benson 1995; Waters et al 1995; Cummings and Clements, 1990; Frick 1985; Lieberman and Michaels, 1962; Coster 1986, Kagan et al, 1988). We havetherefore analyzed a database of the speech of ATCs working under simulated traffic-densities obtained by Dr. O. Veronika Prinzo (Civil Aeronautical Institute, Oklahoma City, Oklahoma) to establish whether previously identified factors show a reliable relationship to ATC simulated workload.

2. Background

Human speech production results from the activity of three functionally distinct systems: (1) the subglottal lungs, (2) the larynx, and (3) the supralaryngeal airway--the supralaryngeal "vocal tract" (SVT). The acoustic consequences of the physiology of these systems have been studied since the early 19th century when Muller (1848) formulated what has come to be known as the "source-filter" theory of speech production. Muller noted that the outward flow of air form the lungs usually provides the power for speech production. If the human auditory system were capable of perceiving acoustic energy at extremely low frequencies, we would "hear" the expiratory airflow. However, the acoustic energy present in the outward flow of air from the lungs is inaudible. The "sources" of acoustic energy for speech are generated by modulating the outward, expiratory flow of air.

Two fundamentally different sources enter into the production of human speech—periodic phonation and turbulent noise sources provide the acoustic energy for speech. Phonation is the result of the activity of the larynx. The vocal folds of the larynx, which are extremely complex structures, move inwards and outwards, converting the steady flow of air flowing outwards from the lungs into a series of "puffs" of air. Both the basic rate and the detailed airflow through the phonating larynx can be modulated by adjusting the tensions of various laryngeal muscles and the alveolar air pressure. The fundamental frequency of phonation (F0) is by definition the rate at which the vocal folds open and close. The perceptual response of human listeners to F0 is the perceived pitch of a speaker's voice. Young children, for example, generally have high F0s during phonation; their voices therefore are "high pitched." Acoustic energy occurs during phonation at the F0 and at the harmonics of the F0. For example, if F0 is 100 Hz, energy can occur at 200 Hz, 300 Hz, etc. The amplitude of the harmonics typically decreases as frequency increases for the phonatory patterns typical of human speech. During the course of speech production, speakers constantly modify the fundamental frequency of phonation for linguistic ends. Distinctions in dialect as well as semantic distinctions can be transmitted by deliberate modifications of the fundamental frequency contour of an utterance. In English, for example, yes-no questions are usually signaled by a rise in F0 at the end of a sentence, stressed words by local peaks in F0 (Lieberman, 1967).

Noise sources are aperiodic and tend to have acoustic energy evenly distributed across all frequencies. Noise sources can be generated at constrictions along the airway leading out from the trachea when the airflow becomes turbulent. Noise can be generated at the larynx by forcing air through the partly abducted vocal cords as, for example, at the start of the word 'hat'. Noise can also be generated by forcing air through constrictions in the SVT. For example, the constriction formed in the mouth when the tongue blade is raised close to the hard palate in the initial consonant of the word 'shoe' generates the noise source of the initial consonant. Momentary bursts of noise excitation typically occur on the release of stop consonants such as [p] when the lips open, at the start of the word 'pig'. The burst is momentary because the turbulent noise abruptly ceases as the airflow changes from turbulent to laminar flow as the lips open wide.

The time interval between the burst of a stop consonant and the onset of phonation of the following vowel is the voice onset time (VOT). VOT differentiates English "voiced stop" consonants like [b], [d], and [g] from their unvoiced counterparts [p], [t], and [k], respectively. In order to produce a [b], a speaker must initiate phonation soon after opening the lips (within about 20ms.) to release the pressure in the vocal tract. In contrast, phonation is delayed for 40 ms. or more after lip opening in a [p]. Similar timing distinctions differentiate [d]s from [t]s and [g]s from [k]s. Figure 1 shows the waveforms for a [b] and a [p] produced by the same speaker, where the lip opening (identified by a visible burst) and the onset of phonation (evidenced by periodicity in the waveform) have been marked. The time delay between the marks is the VOT. Normally, speakers of English and many other languages maintain the distinctions between voiced and unvoiced stop consonants by keeping the VOT regions of the two separated by at least 20 ms.

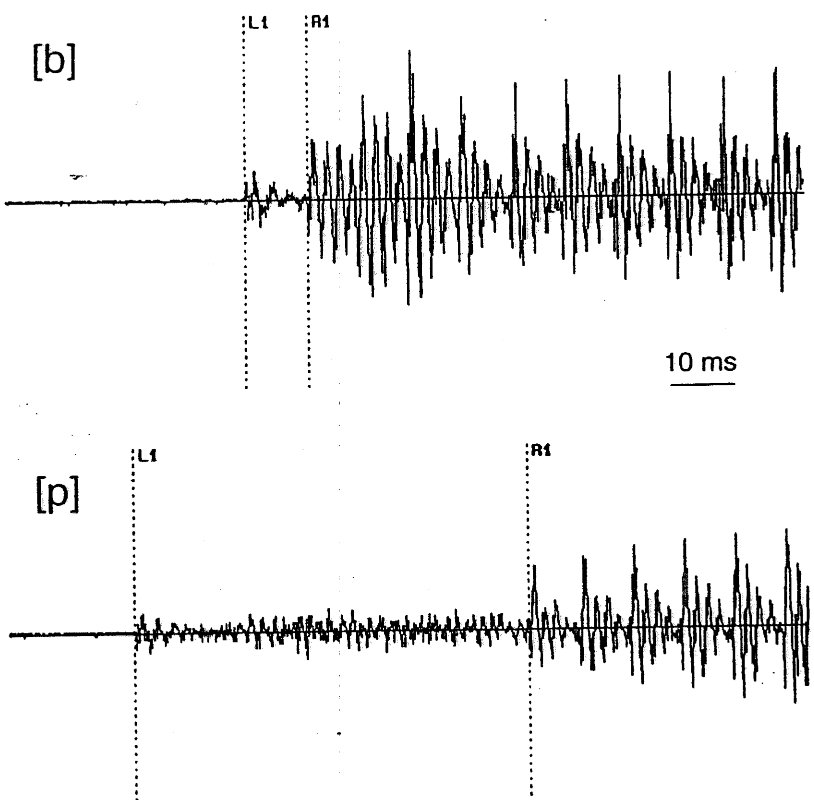


Figure 1. Speech waveform segments corresponding to a [b] and a [p] spoken by the same speaker under identical conditions. Cursors have been placed at the onset of the burst that was caused by opening the lips (L1) and at the onset of periodicity that indicates vocal fold vibration (R1). The marked interval, Voice Onset Timing (VOT) is used by speakers and listeners to differentiate the two types of consonants in word-initial positions.

3. Measures of Interest

We selected four primary measures of interest: speaking rate, hesitation, F0, and VOT.

Speaking rate (syllables/second) might covary with workload in either of two directions. It is possible that because an increase in work load requires an increase in the number of communications in a fixed amount of time, speaking rate would be increased in order to "squeeze in" more information in a given time period. Conversely, it might be the case that as workload increases, speaking rate decreases. It has been shown that verbal 'hesitation' increases with task difficulty and with the quality of a cognitive solution to a given task (Eisler 1968). It has further been shown that there is an inverse relation between the amount of hesitation and speaking rate (Eisler 1968). Assuming that increasing workload is equivalent to increasing task difficulty, speaking rate may decrease as workload increases.

In light of Eisler's findings, hesitation was determined to be a measure of interest. Eisler established that in general, 40% - 50% of speech is actually silence; that is, speech is not the continuous flow of sound indicated by our perception. The silence in connected speech is of three types: 1) the discontinuity of phonation which occurs in articulatory shifts, e.g., when two stop consonants follow each other; 2) discontinuity of phonation attributable to hesitation; and 3) the gap in speech required for inhalation. The second type of silence, hesitation pauses, is associated with complexity of general planning, task difficulty, and the quality of a cognitive solution. Eisler and her colleagues showed that a person making reasoned responses to a question will have longer type 2 pauses than a person responding in an automatic, non-engaged manners. Pause duration thus can reflect "thinking time." Therefore, increased hesitation may be due to two distinct processes, an increase in pause frequency or an increase in pause duration. We therefore used two measures to examine hesitation: number of pauses per word and average duration of pauses per word. Again, if an increase workload results in an increase in cognitive load or thinking time, hesitation, as reflected in either measure may increase.

Although it is generally accepted that fundamental frequency (F0) is affected by physiological and psychological stress, there is conflicting evidence as to which specific properties of F0 are involved (e.g., Lieberman et al, 1995; Absil et al, 1995; Benson 1995; Waters et al 1995; Cummings and Clements, 1990; Frick 1985; Lieberman 1963). We selected a global and a local measure: the overall pitch contour of an utterance, and the pitch period of the highest amplitude portion of the second vowel in the word 'approach' when it occurred utterance-finally.

VOT is an objective, acoustic measures of speech production that reflects a subject's ability to precisely sequence the maneuvers of the tongue, lips, velum and larynx that are necessary to produce human speech. Studies of Broca's aphasia (Blumstein et al, 1980; Baum et al, 1990), Parkinson's Disease (Lieberman et al, 1992) and mountain climbers breathing low-oxygen-content air in the course of an ascent of Mount Everest (Lieberman et al, 1995) show that control of VOT deteriorates. In these cases, abnormal VOT production is correlated with decrements in reasoning and sentence comprehension. As such, it has been suggested that VOT production may be used as an index of cognitive functioning. Accordingly, we measured the VOT of the word-initial voiceless velar stop [k] and the word-medial voiceless alveolar stop [t] from the word 'contact' when it occurred utterance-finally in the phrase 'contact approach'.

4. Methods

Two approaches were taken in the speech analysis, narrow and broad. First, a detailed examination of the speech of a single subject, Dallas Fort Worth Subject 1, was performed. Measures were taken from all utterances produced by this subject in two scenarios, Feeder East Heavy and Feeder West Heavy. Second, a subset of utterances produced by the remaining subjects were analyzed to assess the generalizability of Subject 1 results. Utterances were selected to achieve maximal workload contrast for each subject. For each subject, ten utterances were identified and digitized. These are the five lowest workload utterances from the Light version of the scenario simulating that subject's normal work station and the five highest workload utterances from the Heavy version of the scenario simulating a work station unfamiliar to the subject. All speech signals were sampled at 16 bits quantization at 20,000 samples per second; the digitized signal was stored in audio files.

The analysis was performed using the interactive BLISS speech analysis system developed by John Mertus ((Lieberman and Blumstein, 1988). The BLISS system permits trained operators to monitor and modify analysis parameters at virtually all stages of analysis, thereby minimizing artifacts that otherwise can be introduced by most commercially available speech analysis software. The BLISS system allows operators to view the waveform and position 4 independent sets of "cursors," e.g., left cursor L0 and right cursor R0, on the waveform. The operator can listen to the waveform delineated by any set of cursors; the sectioned waveform can be transferred to another file, "spliced" to any other file, reduplicated, scaled up or down in amplitude, inverted, etc. Figure 2 illustrates some of the features of the BLISS system. The amplitude of the speech signal at the onset of the word is displayed on the ordinate as a function of time which is plotted with respect to the abscissa.

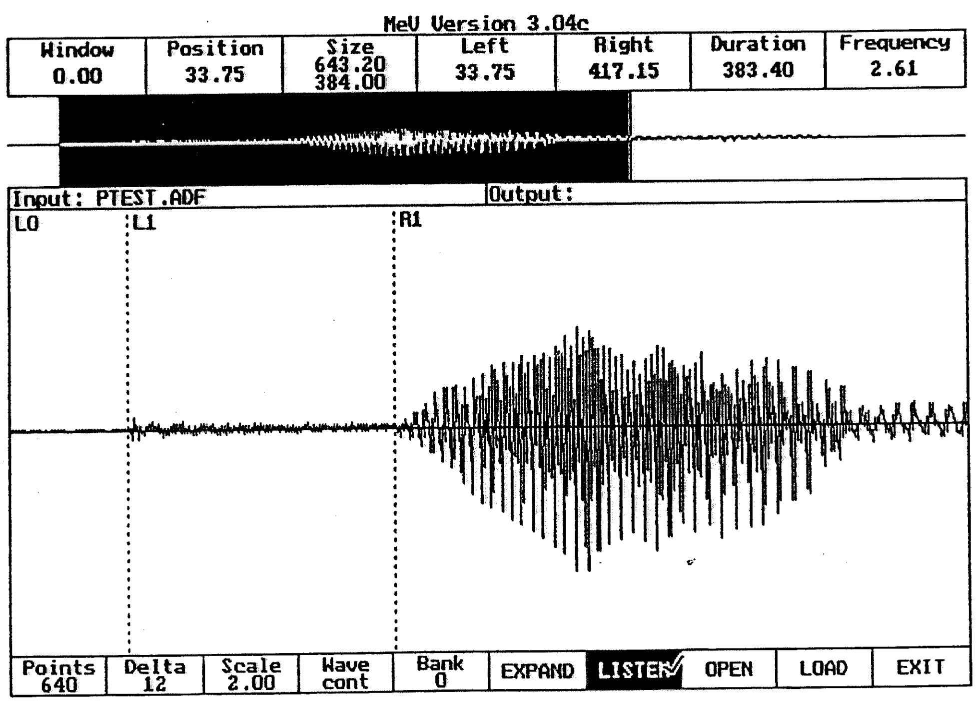


Figure 2. BLISS system display showing waveform for the word 'pig'. Cursors L1 and R1 mark VOT of word initial consonant [p].

The upper part of Figure 2 shows various aspects of the BLISS system "header" identifying the stored audio file, the cursor positions, and the waveform of the complete file, i. e., the word 'pig'. The lower boxes control a number of parameters of the BLISS system by means of a mouse and display the chosen parameter values. "Points" indicates the number of points that are displayed on the screen; they can be varied from 32 to 2480, allowing the operator to view and manipulate the signal with different temporal resolution. The "Delta" command instructs the system to display every Xth data point, compressing the signal. "Wave cont" is a switch that can be set to display individual data points, or as in the display of Figure 2, interpolate between data points. The "Bank" box allows one of four sets of cursors to be displayed and moved. "Expand" transfers the reverse-field (black background) display from the upper waveform display to the full screen. "Listen" allows the operator to listen to the section between any set of cursors or the total waveform displayed above. The "open" and "load" boxes are used to open new files and to transfer data to these files, e. g., the waveform between any set of displayed cursors.

For all subjects, three measures were generated from the utterances selected as described above. 1) **Speaking rate** (syllables/second) was computed from number of syllables per utterance and utterance duration in ms. Because ellisions and contractions (e.g., 'merican' for 'American') were common, only syllables actually uttered, as determined by listening to the speech sample and by visual examination of the waveform, were counted, rather than number of syllables prescribed by standard English pronunciation. Utterance duration was measured by placing cursors at the onset and offset of speech as determined by visual examination of the waveform and by listening to the speech sample. 2) **Pause frequency** (number of pauses/number of words) was computed from the number of pauses per utterance and the number of words per utterance. The pauses in speech are normally of too short a duration to be auditorily perceptible. Thus pauses were identified by visual examination of the waveform. A pause was defined to be a 'flat' portion of the waveform greater than 25 ms. Some articulatory gestures of speech (e.g., stop consonants) necessarily result in brief periods of silence. Using a 25 ms. lower bound excludes these articulatory factors. 3) **Pause duration** (duration of pauses/number of words). The duration in ms. of each identified pause was measured by placing cursors at the onset and offset of silence as indicated by flattening of the waveform.

The comprehensive analysis of Subject 1's speech included three additional measures. For each utterance, an **F0 track**, or pitch track was computed for the entire utterance. Pitch analysis was done by the Short Term Autocorrelation algorithm. Fundamental frequency is the lowest harmonic in the Fourier decomposition of a complex waveform. The Autocorrelation method extracts this harmonic from the waveform (Lieberman and Blumstein, 1986). The resulting pitch tracks were analyzed using an interleavings and offsets method, in which individual pitch tracks are interleaved and the spread, or offset, is assessed.

Additional measurements were performed on the subset of utterances which included the words 'contact approach' in utterance-final position (26% of utterances). This data set was chosen to minimize variation attributed to context, both lexical and phrasal. **VOTs** were measured for the voiceless velar stop [k] at the beginning of the word 'contact' and the voiceless alveolar stop [t] at the beginning of the syllable '-tact'. Cursors were placed at the onset of the burst produced at the release of the each stop consonant and at the onset of phonation, by means of both visual inspection of the waveform and by listening to marked portions of the signal. The duration of the **pitch period** of the highest amplitude portion of the vowel in the syllable '-proach' was measured by placing cursors on two successive peaks of the waveform.

5. Results

5.1 Narrow Focus, Dallas Fort Worth Subject 1

Figures 3-14 show average speaking rate, average pause frequency, average pause duration, velar VOT, alveolar VOT, and pitch period as a function of workload for Subject 1 in Dallas Fort Worth East and West Heavy scenarios. Regression analyses performed for each measure showed no significant correlation between any measure and workload. Speaking rate east: r=0.0383, p=0.6417; Speaking rate west: r=0.0950,

Figure 3. Average speaking rate(syllables/second) as a function of workload for Subject 1 in the Dallas Fort Worth East Heavy Scenario.

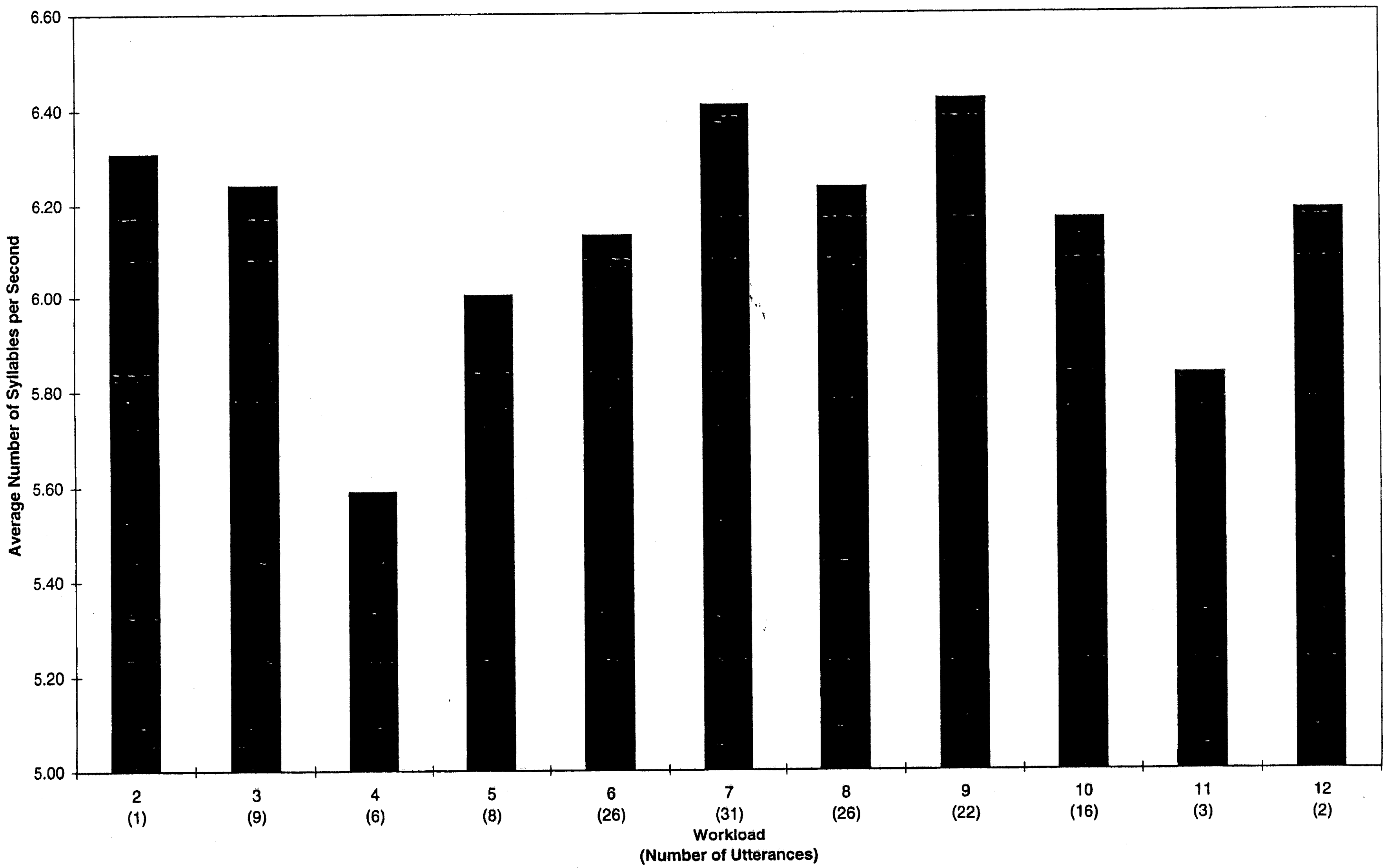


Figure 4. Average speaking rate (syllables/second) as a function of workload for Subject 1 in the Dallas Fort Worth West Heavy Scenario.

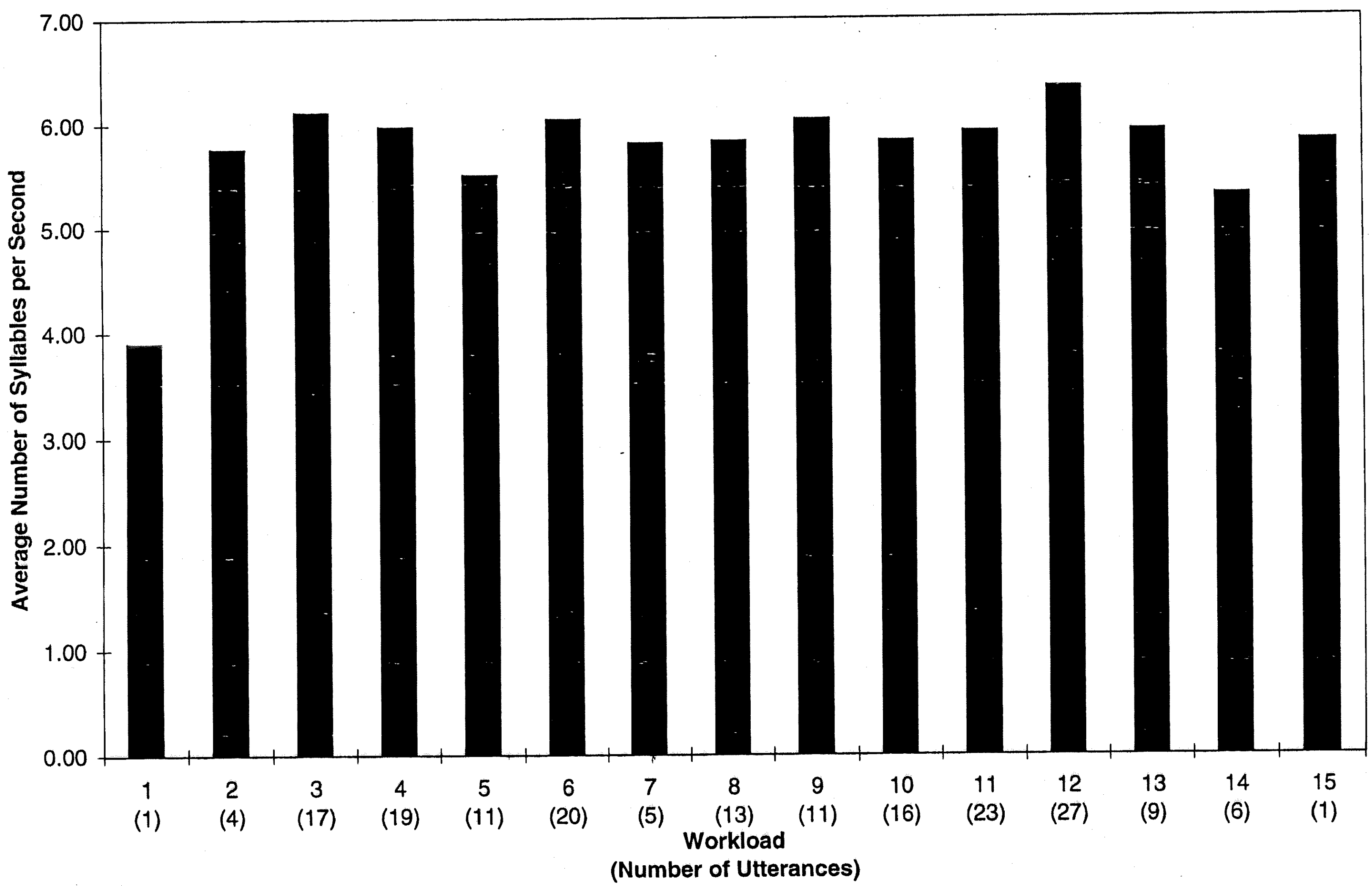


Figure 5. Average pause frequency (number of pauses/number of words) as a function of workload for Subject 1 in the Dallas Fort Worth East Heavy Scenario.

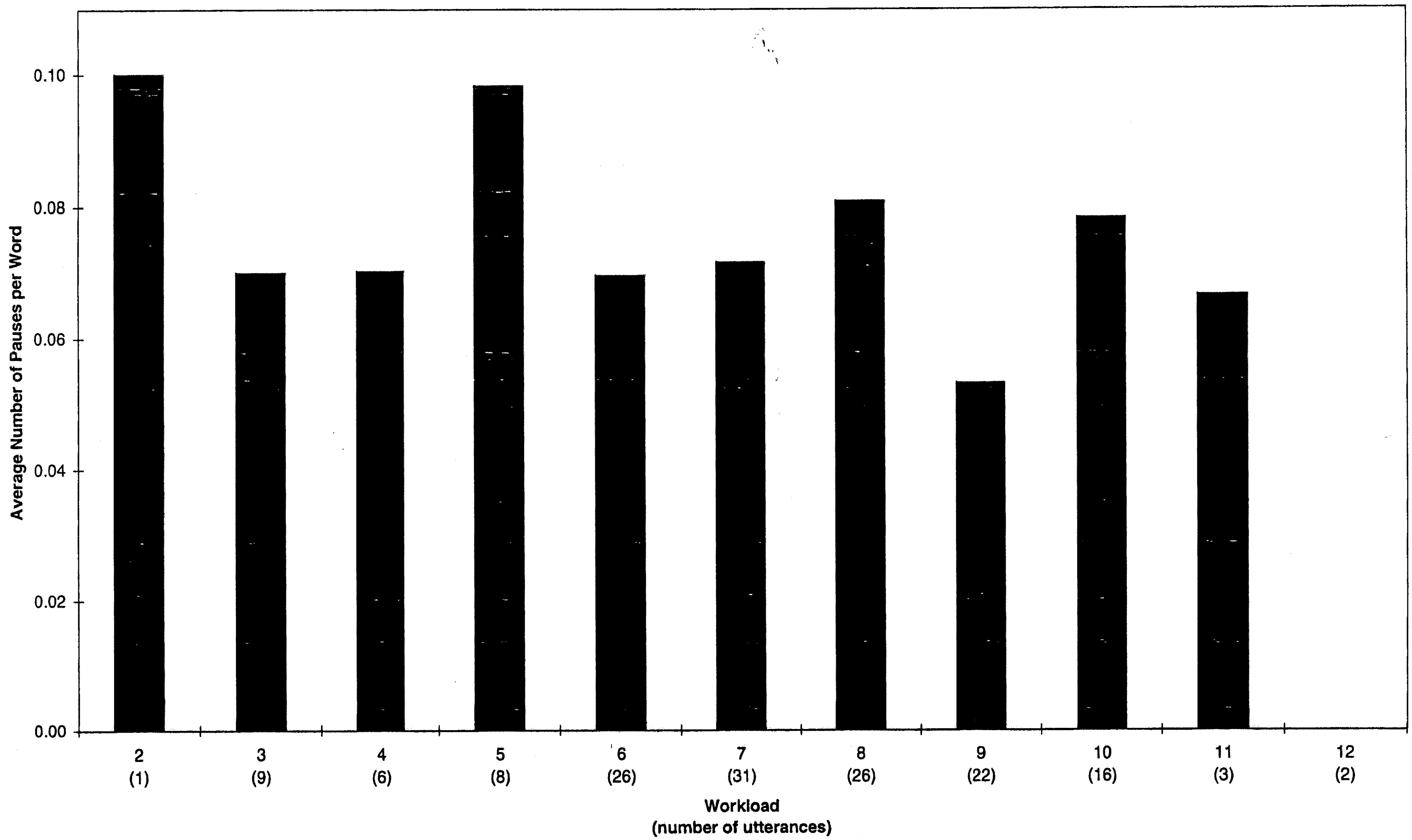


Figure 6. Average pause frequency (number of pauses/number of words) as a function of workload for Subject 1 in the Dallas Fort Worth West Heavy Scenario.

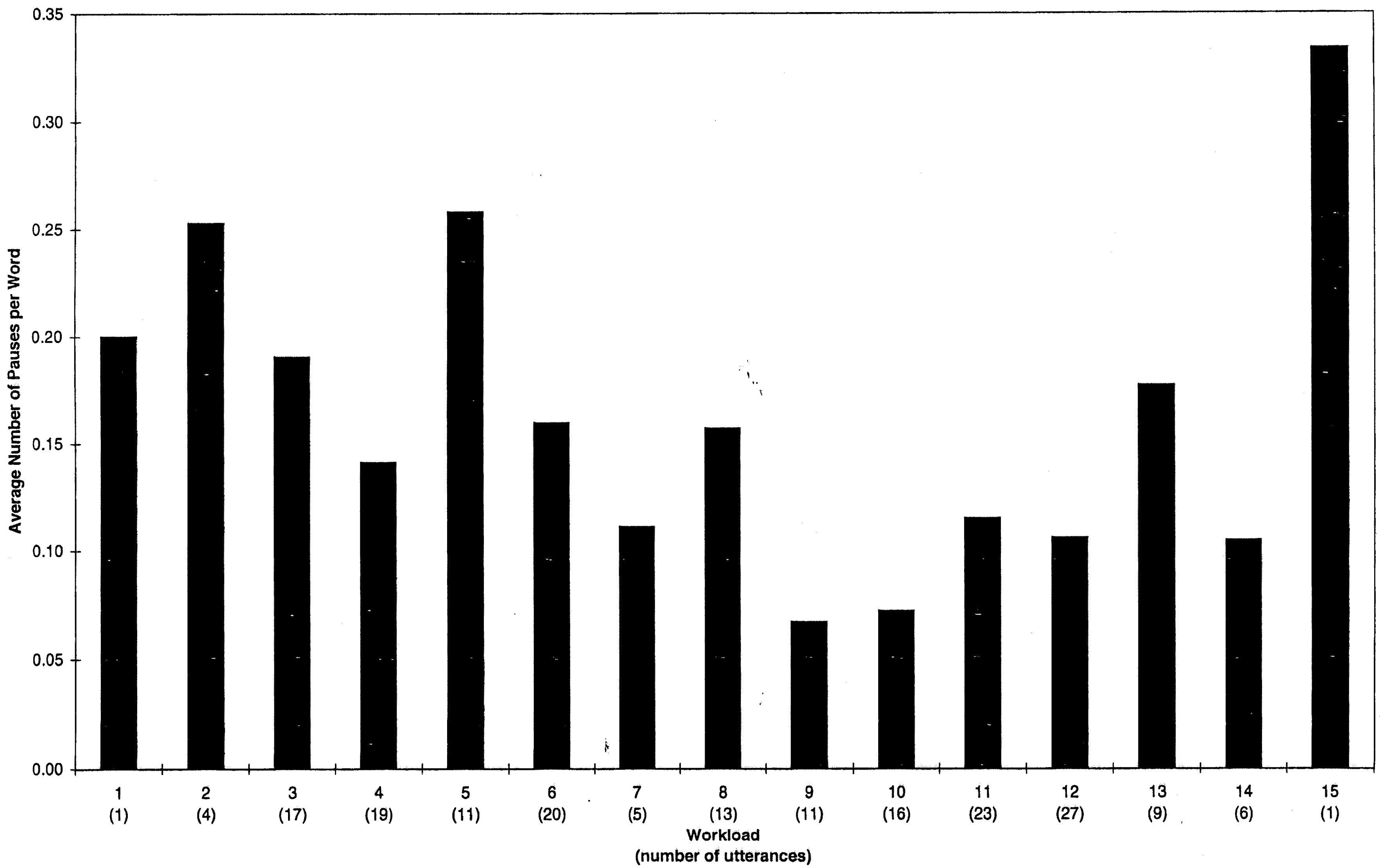


Figure 7. Average pause duration (duration of pauses/number of words) as a function of workload for Subject 1 in the Dallas Fort Worth East Heavy Scenario.

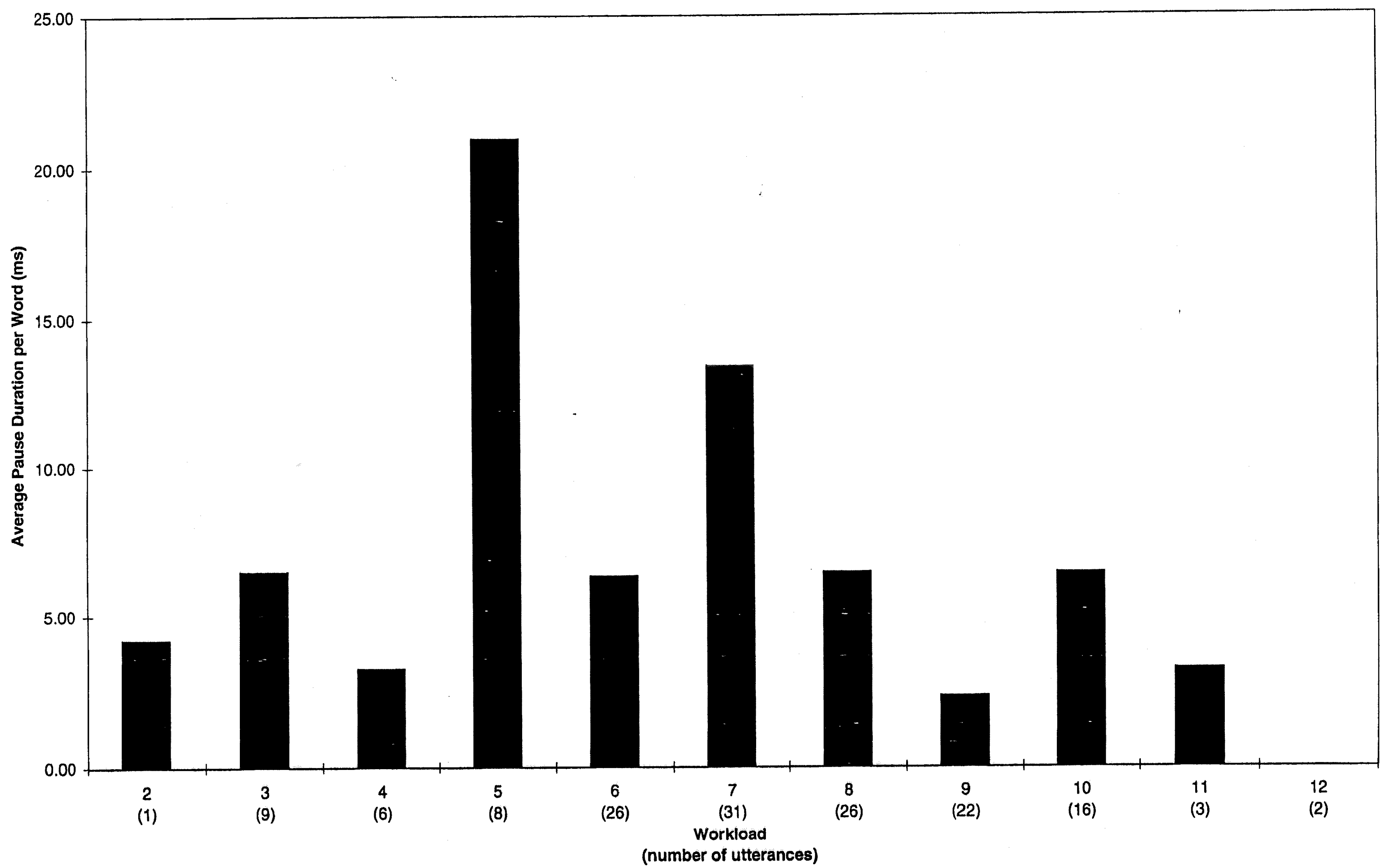


Figure 8. Average pause duration (duration of pauses/number of words) as a function of workload for Subject 1 in the Dallas Fort Worth West Heavy Scenario.

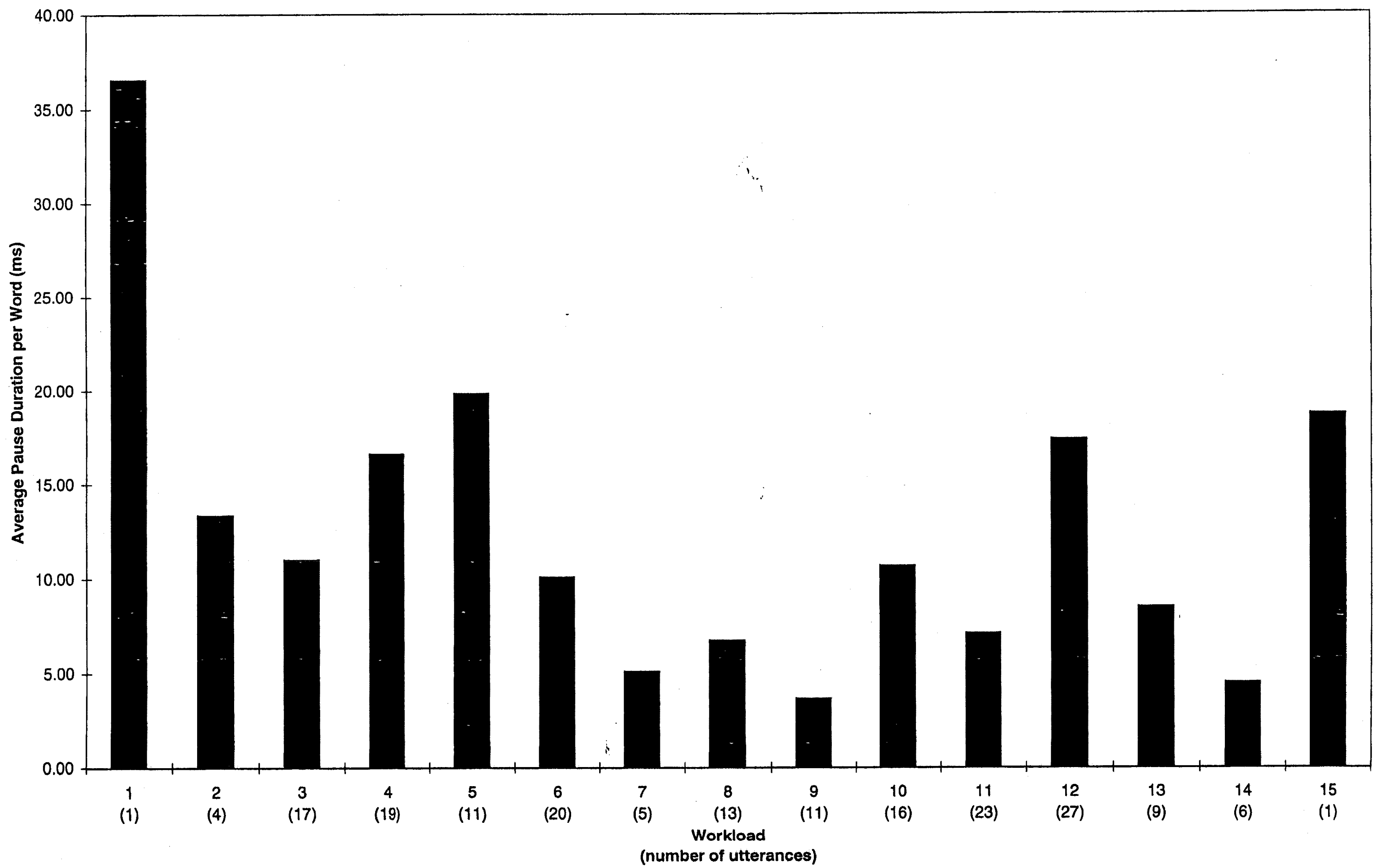


Figure 9. Average VOT of [k] as a function of workload for Subject 1 in the Dallas Fort Worth East Heavy Scenario.

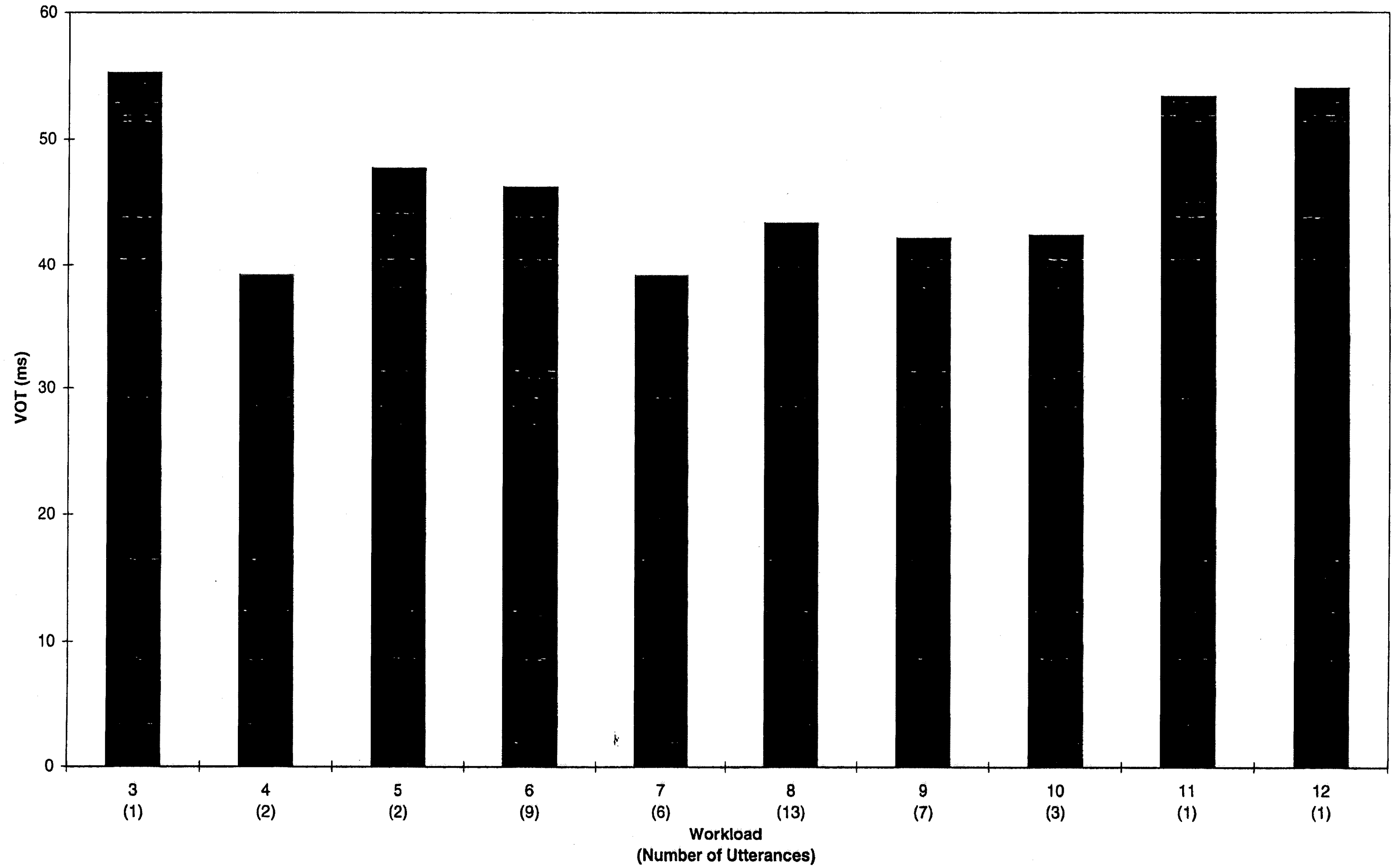


Figure 10. Average VOT of [k] as a function of workload for Subject 1 in the Dallas Fort Worth West Heavy Scenario.

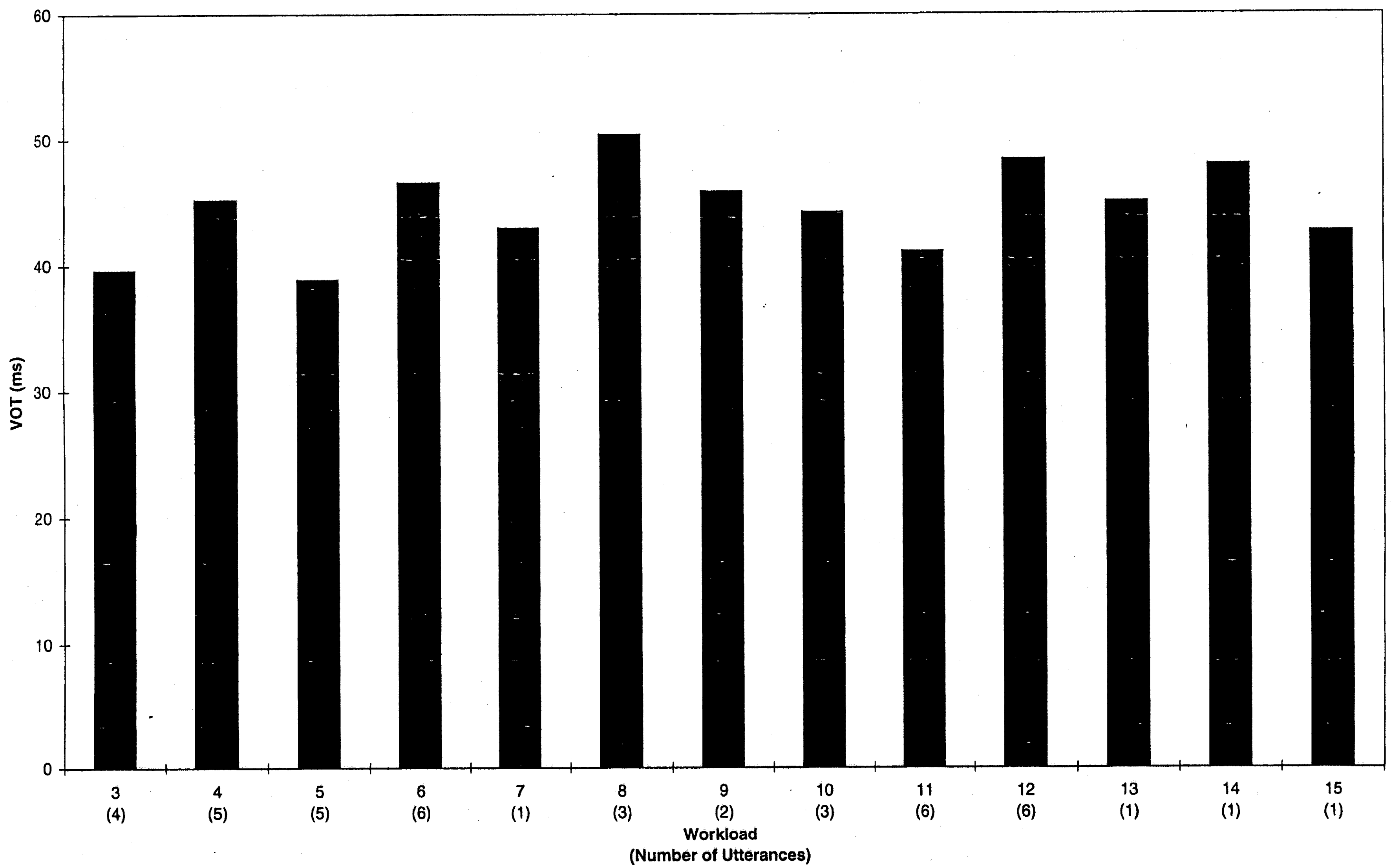


Figure 11. Average VOT of [t] as a function of workload for Subject 1 in the Dallas Fort Worth East Heavy Scenario.

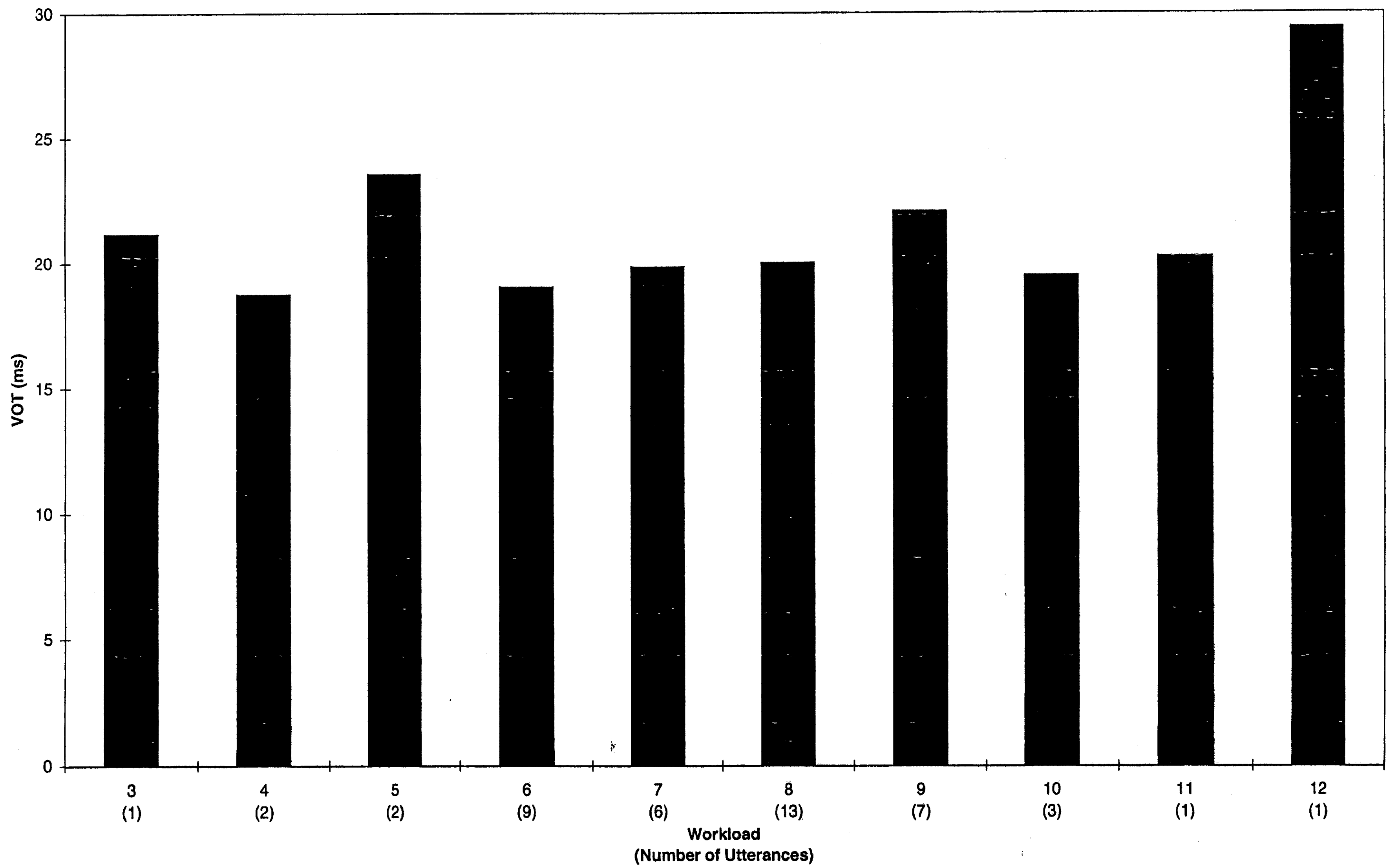


Figure 12. Average VOT of [t] as a function of workload for Subject 1 in the Dallas Fort Worth West Heavy Scenario.

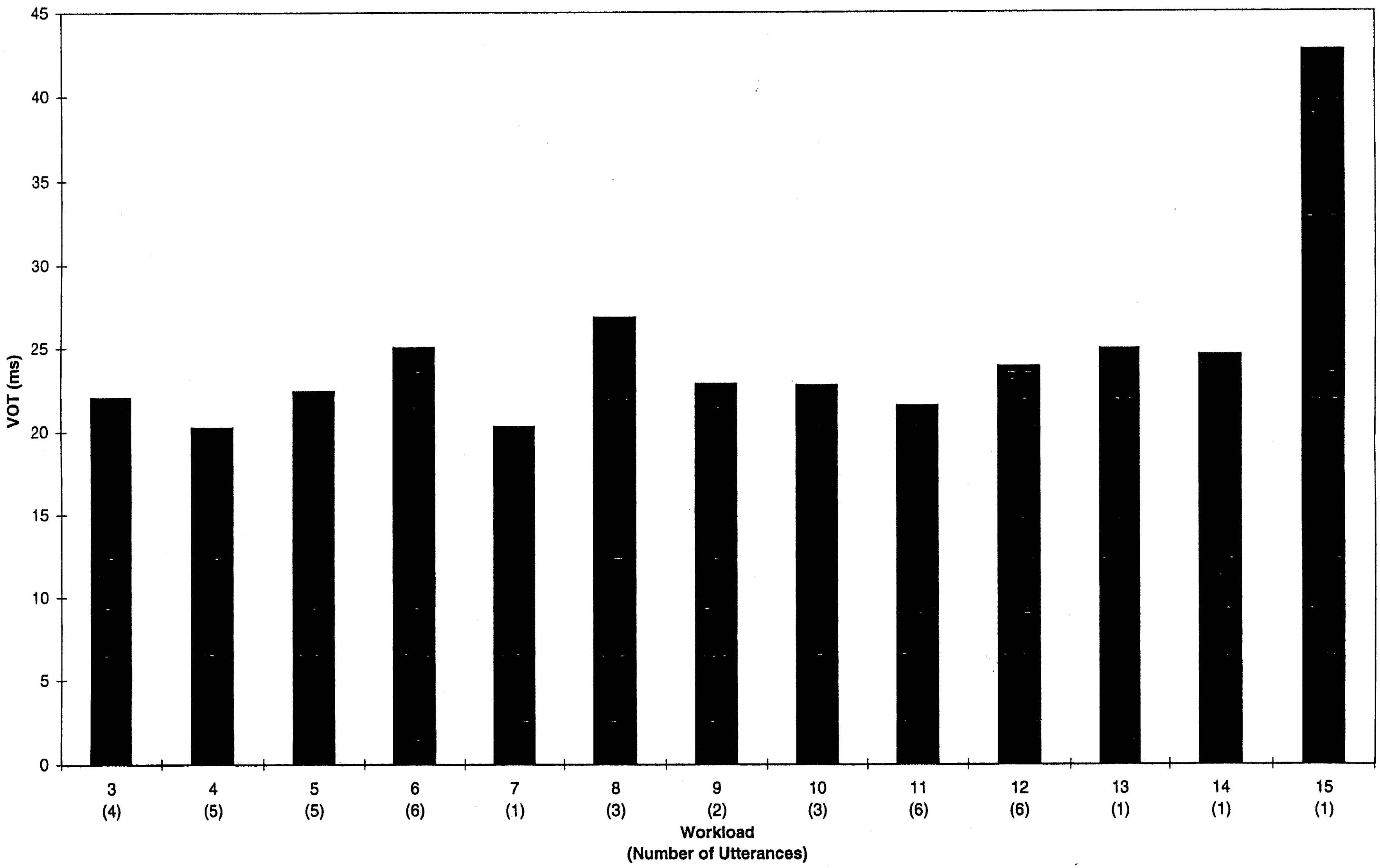


Figure 13. Average pitch period from phrase final syllable 'proach' as a function of workload for Subject 1 in the Dallas Fort Worth East Heavy Scenario.

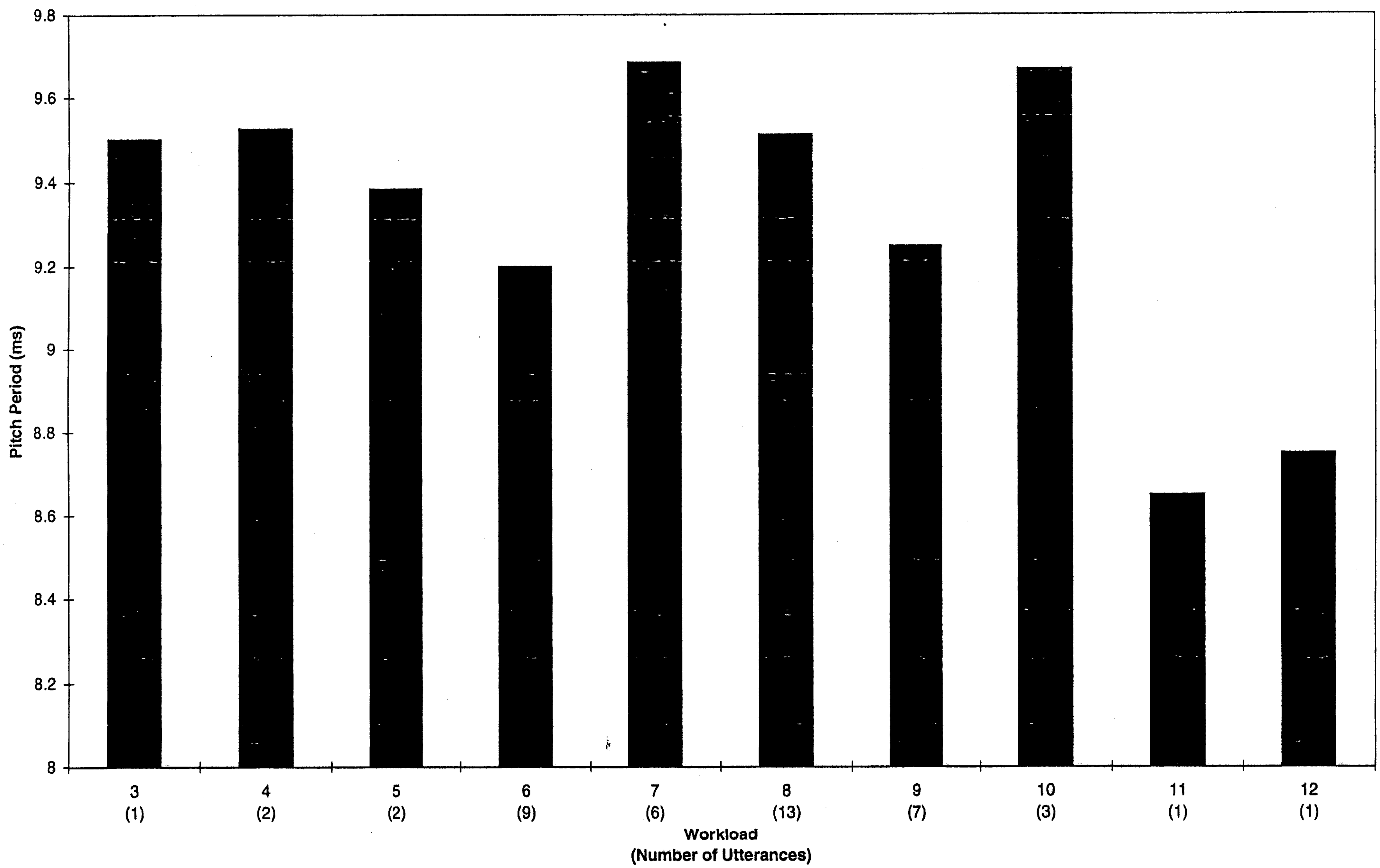
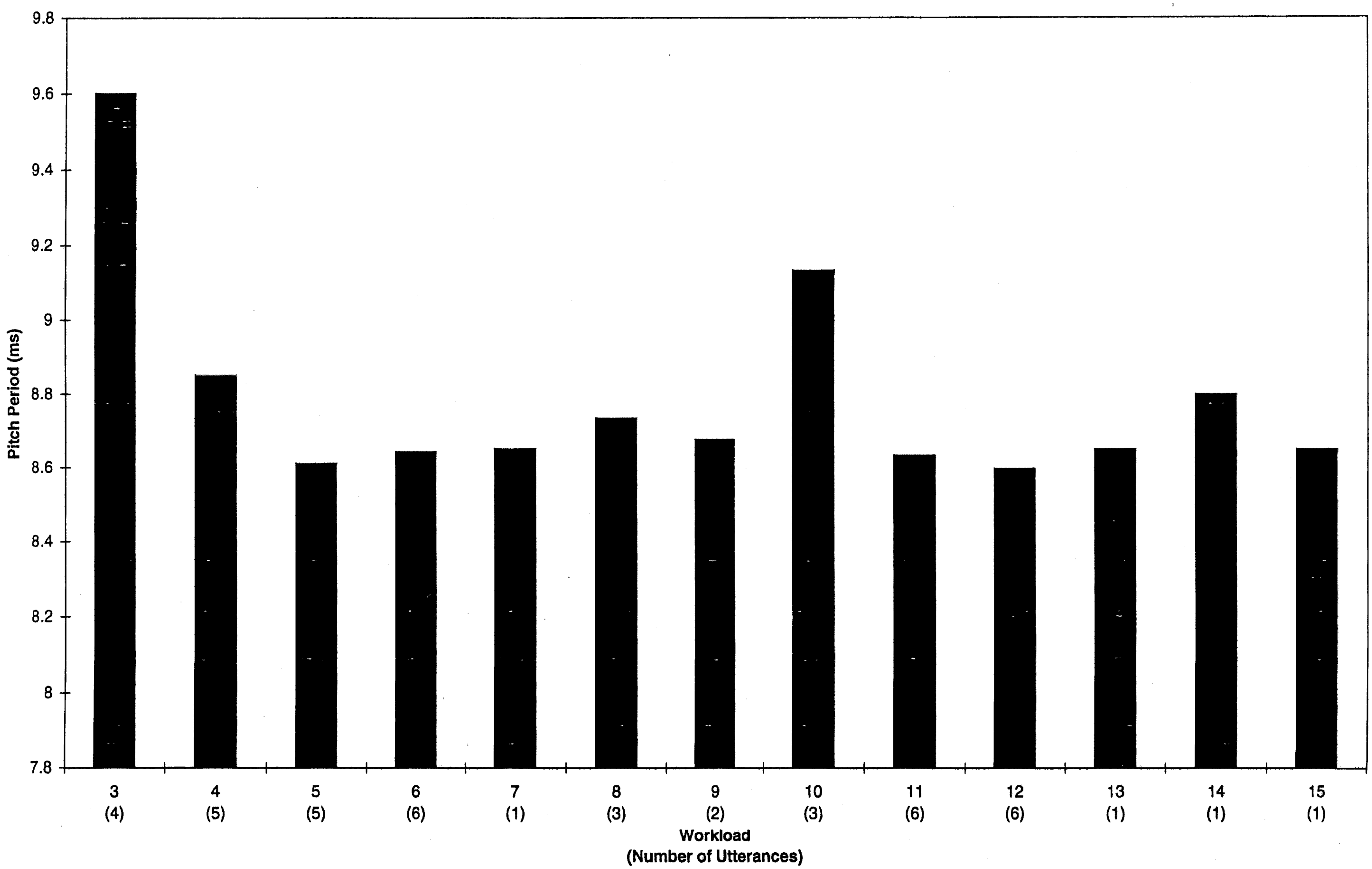


Figure 14. Average pitch period from phrase final syllable 'proach' as a function of workload for Subject 1 in the Dallas Fort Worth West Heavy Scenario.



p=0.2009; Pause frequency east: r=-0.0575, p=0.4843; Pause frequency west: r=-0.1702, p=0.0213; Pause duration east: r=-0.0758, p=0.3565; Pause duration west; r=-0.0599, p=0.4209; Velar VOT east: r=-0.581, p=0.7045; Velar VOT west: r=0.1125, p=0.4671; Alveolar VOT east: r=0.1058, p=0.4891; Alveolar VOT west: r=0.2729, p=0.0731; Pitch period east: r=-0.1485, p=0.3302; Pitch period west: r=-0.2729, p=0.0731. Examination of the F0 tracks revealed variations in contour dynamics within the range noted in previous studies of single speakers (Atkinson, 1973; Lieberman et al, 1986) and therefore were concluded to be nonsignificant.

5.2 Broad Focus, Dallas Fort Worth Subjects 2 - 12

5.2.1 Speaking Rate

Figure 15 shows average speaking rate (SR) for each subject in the Light and the Heavy conditions. Overall, there appears to be a trend towards faster speech (more syllables per second) in the Light than in the Heavy condition. Seven of the 11 subjects show this pattern, 3 show the reverse pattern, and one subject's SR does not change across conditions. Statistical comparisons performed for each subject showed significant differences for two subjects: Subject 8's SR was significantly faster in the Heavy condition ($p = 0.02$) and Subject 12's SR was significantly faster in the Light condition ($p = 0.03$).

In order to investigate whether the observable trend towards faster speech in the Light condition was significant on a group basis, normalized SR measures were generated. This was done by computing, for each subject and each SR value, the ratio of that value to the subject's mean SR. A statistical comparison was then performed on the normalized SRs from the Light and Heavy conditions collapsed across subjects. The mean normalized SRs for the Light and Heavy conditions were 1.03 and .097, respectively. This difference was not significant ($p = 0.10$).

5.2.2 Pause Frequency

Figure 16 shows average pause frequency (PF) for each subject in the Light and the Heavy conditions. Again, there appears to be a trend towards more frequent pauses in the Light

Figure 15. Average speaking rate (SR) for each subject in the Light and the Heavy conditions.

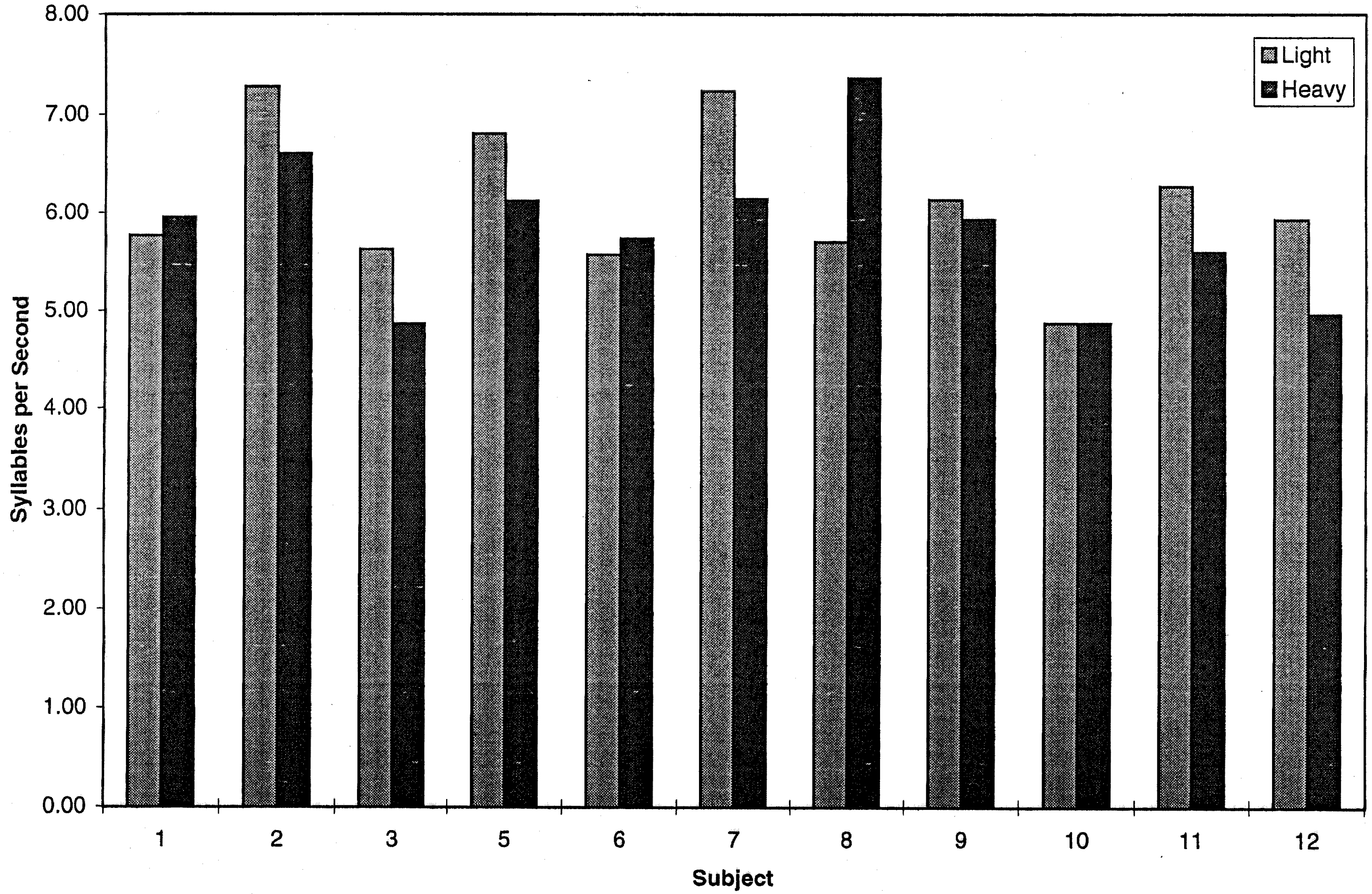
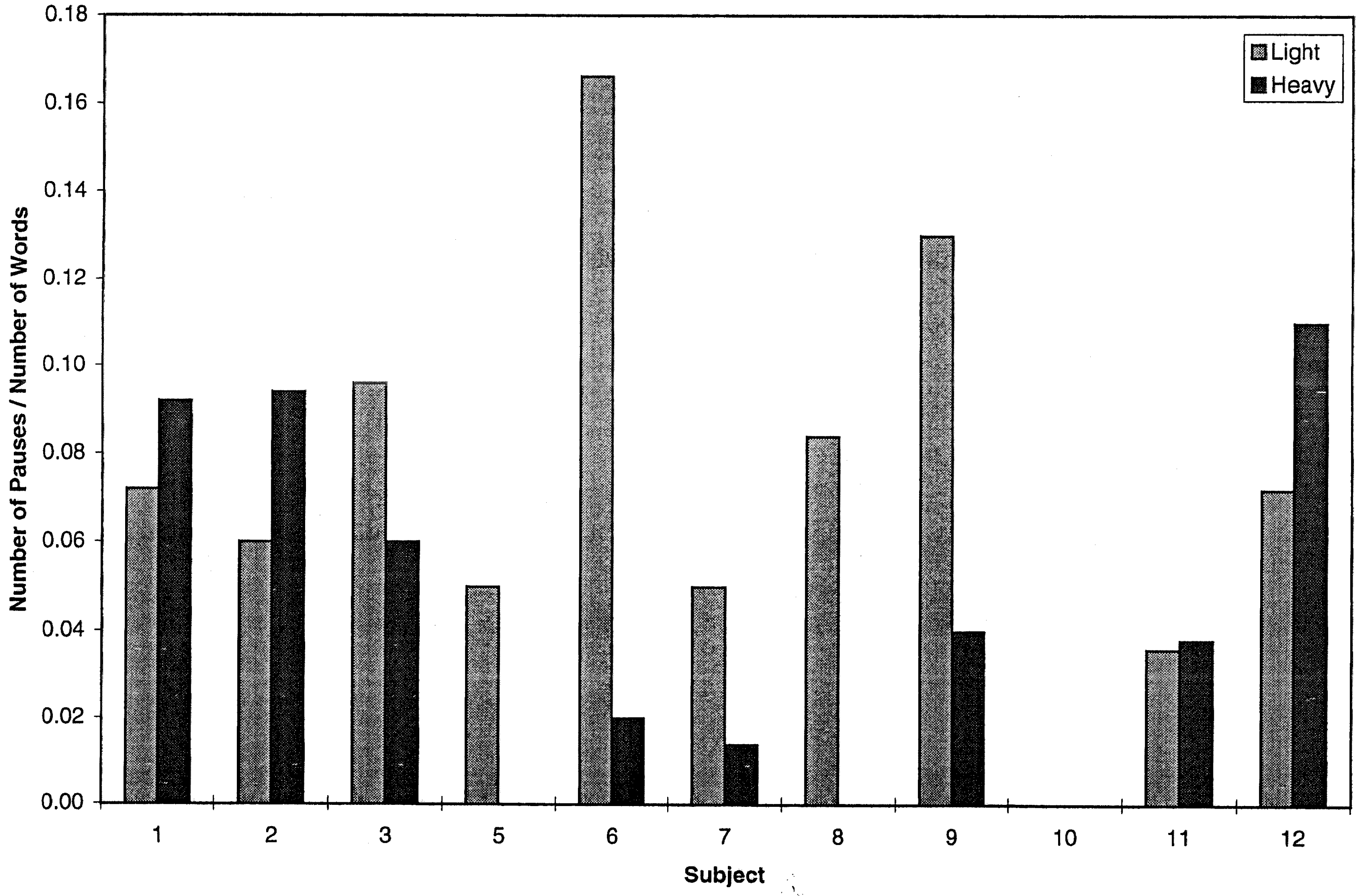


Figure 16. Average Pause Frequency for each subject in the Light and the Heavy conditions.



than in the Heavy condition. Six of the 11 subjects show this pattern, 4 show the reverse pattern, and one subject did not produce any pauses. Statistical comparisons performed for each subject showed a significant difference for one subject: Subject 6's PF was significantly greater in the Light than in the Heavy condition ($p = 0.002$).

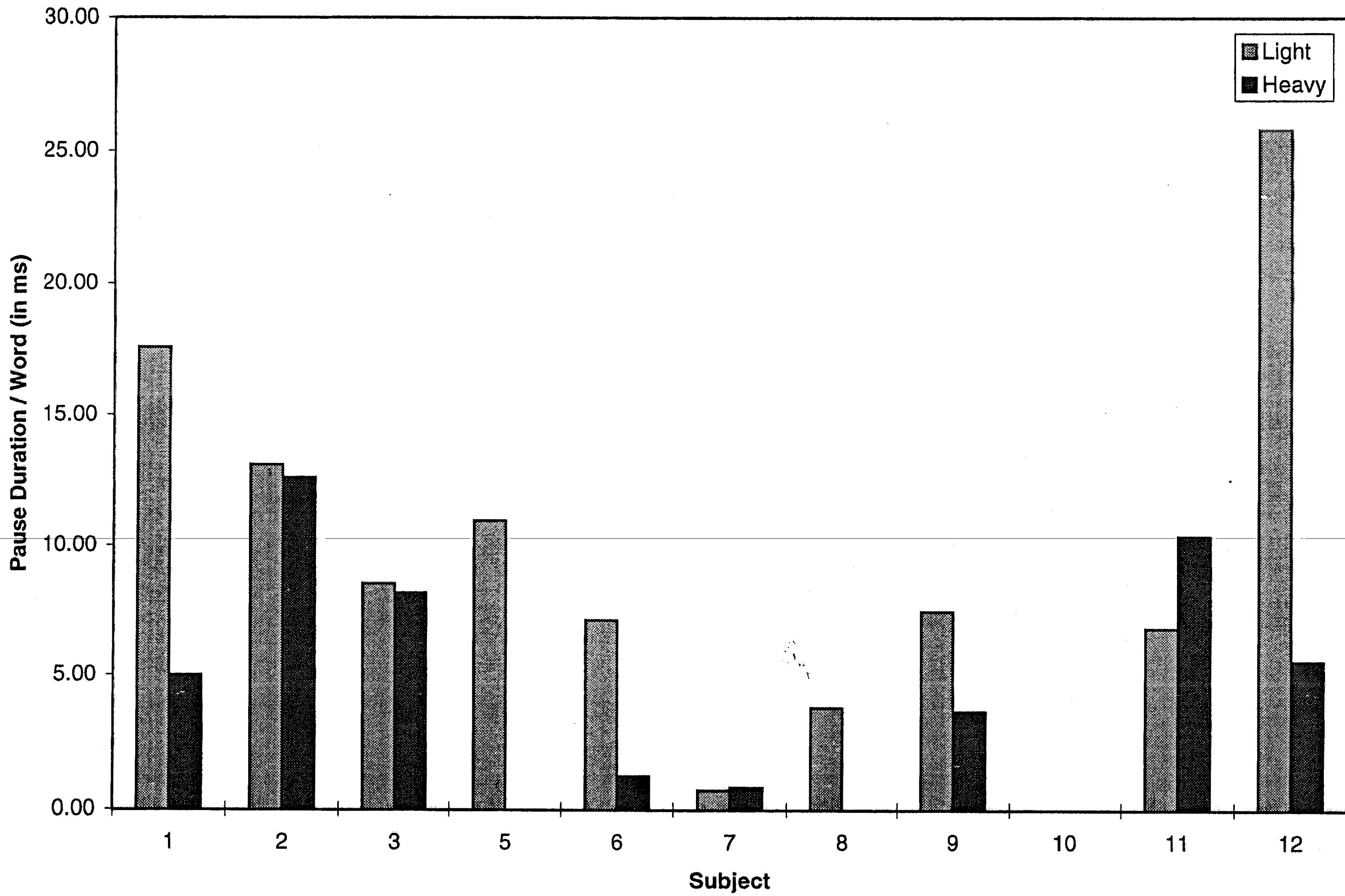
In order to investigate whether the observable trend towards greater pausing in the Light condition was significant on a group basis, normalized PF measures were generated. This was done by computing, for each subject and each PF value, the ratio of that value to the subject's mean PF. A statistical comparison was then performed on the normalized PFs from the Light and Heavy conditions collapsed across subjects. The mean normalized PFs for the Light and Heavy conditions were 1.45 and 0.54, respectively. This difference was significant ($p = 0.005$).

5.2.3 Pause Duration

Figure 17 shows average pause duration (PD) for each subject in the Light and the Heavy conditions. Again, there appears to be a trend towards greater pause duration in the Light than in the Heavy condition. Eight of the 11 subjects show this pattern, 2 show the reverse pattern, and one subject did not produce any pauses. Statistical comparisons performed for each subject showed a significant difference for one subject: Subject 8's PD was significantly greater in the Light than in the Heavy condition ($p = 0.05$).

In order to investigate whether the observable trend towards greater pausing in the Light condition was significant on a group basis, normalized PD measures were generated. This was done by computing, for each subject and each PD value, the ratio of that value to the subject's mean PD. A statistical comparison was then performed on the normalized PDs from the Light and Heavy conditions collapsed across subjects. The mean normalized PDs for the Light and Heavy conditions were 1.46 and 0.54, respectively. This difference was significant ($p = 0.005$).

Figure 17. Average Pause Duration for each subject in the Light and the Heavy conditions



6. Discussion

The results presented here suggest two points of interest. First, as a group, the ATCs in this simulation have a tendency both to pause more frequently and to pause for greater durations in a light work load situation than in a heavier workload situation. From this result, it is possible to infer that the type of 'hesitation' produced by these ATCs is not associated with factors such as task difficulty, as described above in Section 3. Instead, these data may reflect one or both of the following possibilities: 1) Task load is light, therefore the ATC slows down his/her speech and pauses more often and longer. 2) The task load being light, the ATC is able to attend to the task in hand using a "cognitive" rather than an "automatic" response mode. This second possibility is more likely in light of previous studies of pause duration and frequency (Eisler, 1968) and given the lack of a significant reduction in speaking rate as described above.

It is especially interesting that both measures of hesitation increased. The measures used are in principle independent of one another. That is, because pause duration is averaged across all pauses in an utterance, there is no *a priori* reason to suppose that an increase in the number of pauses would be associated with an increase in the length of those pauses. And in fact, it was often the case that a subject who showed the dominant trend on one of these measures showed the opposite of the dominant trend on the other measure. Only 6 of the 11 subjects showed both trends.

The light scenario utterances were on average slightly longer in duration (3218 ms vs. 3027 ms) and slightly greater in number of syllables (18.6 vs. 16.6). Although these differences were not significant, it may be the case that 'more speech' provides more opportunity both for more pauses and for longer pauses. However, the occurrence of both longer Type 2 normalized pause durations and normalized pause frequency are consistent with the ATCs responding in a more cognitive mode in the Light condition.

The second point of interest is that while there were strong group effects for the two measures of hesitation, these effects were rarely significant on an individual basis. Further, none of the three measures showed trends in the same direction for all subjects, regardless of the size of the effects. This was especially true for speaking rate, for which two subjects showed significant differences on opposite directions.

The importance of this fact is highlighted by the results from our more in depth analysis of the speech of Subject 1. Despite our analyzing a large number of utterances (more than 300), which should bring out even weak effects, we found no correlation between any measure and workload. We concluded our progress report by suggesting that Subject 1 may not be representative of the pool of subjects. The group analysis reported here shows this to be both true and false: it is true in that some subjects showed significant differences between the light and heavy conditions for some measures; it is false in that most subjects did not show such effects on an individual basis. Subject 1 is therefore indicative of the large amount of between subject variability documented here.

In conclusion, it appears that hesitation in speech is a potential indicator of workload. Despite its highly speaker-dependent nature, it may be a useful indicator of an ATC's responding in a cognitive rather than in an automatic mode. The exhaustive data collected by Eisler and her colleagues shows that individuals who are devoting fewer cognitive resources to a discussion manifest shorter Type 2 pause durations than people thinking about what they are communicating. The speech of the ATCs in this study, therefore, may reflect a shift between a more cognitive 'thinking' response mode in Light traffic situations where they know that they have more time to respond and a more automatic mode which allows them to respond to the increased pace induced by higher traffic loads. In other words, we may be monitoring the degree to which the ATCs respond by means of reasoned, cognitive rather than automatic, routinized responses. Our research suggests that future investigation of the speech of ATCs should include measures of hesitation as a measure of interest.

REFERENCES

Absil, E., Grammatica, B. Harmegnies, B., Legros, C., Poch, D., and Ruiz, R. (1995) *Time related variabilities in stressed speech under laboratory and real conditions.* Proceedings, ESCA - NATO Tutorial and Research Workshop on Speech under Stress.

Baum, S.R., Blumstein, S.E., Naeser, M.A., and Palumbo, C.L. (1990) *Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech.* Brain and Language. 37:327-338.

Benson, P. (1995) *Analysis of the acoustic correlates of stress from an operational aviation emergency.* Proceedings, ESCA - NATO Tutorial and Research Workshop on Speech under Stress.

Blumstein, S.E., Cooper, W., Goodglass, H. Statlender, H. and Gottleib, J. (1980) *Production deficits in aphasia: a voice-onset time analysis.* Brain and Language 9:153-170.

Coster, W.J. (1986) *Aspects of voice and conversation in behaviorally inhibited and uninhibited children.* Unpublished PhD dissertation, Harvard University.

Cummings, K.E., and Clements, M.A. (1990) *Analysis of glottal waveforms across stress styles.* Proceedings, IEEE ICASSP 369-372.

Eisler, F.G. (1968) Psycholinguistics: Experiments in spontaneous speech. London: Academic Press.

Frick, R.W. (1985) *Communicating emotion: The role of prosodic features.* Psychological Bulletin 97:419-429.

Kagan, J., Reznick, J.S., and Snidman, N. (1988) *Biological bases of childhood shyness.* Science 240:167-171

Kessinger, R. and Blumstein, S.E. (in submission) *Rate of Speech Effects on Voice Onset Time in Thai, French, and English.*

Lieberman, P. (1967) Intonation, perception and language. Cambridge: MIT Press

Lieberman, P. and S. E. Blumstein. (1988) Speech physiology, speech perception, and acoustic phonetics. Cambridge: Cambridge University Press.

Lieberman, P., Kako, E.T., Firedman, J., Tajchman, G., Felldman, L.S., and Jiminez, E.B. (1992) *Speech production, syntax comprehension, and cognitive deficits in Parkinson's disease.* Brain and Language 43:169-189.

Lieberman, P. and Michaels, S.B. (1962) *Some aspects of fundamental frequency, envelope amplitude and the emotional content of speech.* Journal of the Acoustical Society of America 34:922-927.

Lieberman, P., Protopapas, A. and Kanki, B.G. (1995) *Speech production and comprehension deficits on Mt. Everest.* Aviation, Space, and & Environmental Medicine. 66:9 857-869.

Lisker, L. and Abramson, A. S. (1964) *A cross language study of voicing in initial stops: acoustical measurements*. Word 20:384-442.

Muller, J. (1848) The physiology of the senses, voice and muscular motion with the mental faculties. Transl. W. Baly. London: Walton and maberly.

Waters, J., Nunn, S. Gillcrist, B. and VonColln, E. (1995) *The effect of stress on the glottal pulse*. Proceedings, ESCA - NATO Tutorial and Research Workshop on Speech under Stress.